

Statewide Data Warehouse Project

POST PROJECT REVIEW REPORT

Project Overview

In October 2002, the Department of Public Instruction entered into contract with TetraData Corporation to develop and implement a statewide data analysis clearinghouse for all public schools and districts within the State. The project emerged from a growing need to address several policy and programmatic challenges.

- A. The reauthorized federal *Elementary and Secondary Education Act* established clear accountability measures that required districts and the State to collect and analyze data on student achievement and other school-related factors. Additionally, if the State were to exercise its right to amend the model and formula for identifying low-performing schools (Adequate Yearly Progress) based on a research-based cohort-improvement model, a data linkage and analysis tool would be required.
- B. The State's school improvement and professional development requirements demand that schools develop detailed demographic and achievement profiles that require disaggregated information to be oftentimes collected in various databases.
- C. Districts and the State have suffered from an inadequate ability to dissect and analyze data for the purposes of identifying promising or ineffectual instructional or administrative practices.
- D. Districts have acquired data collection applications to collect essential information; however, these applications are limited in their ability to link outside data sources, to conduct sophisticated analysis, or to generate practical reports for study by staff and policy makers.
- E. Seeking Uniformity and an Economy of Scale
 - 1. It is more desirable to establish a means of dissecting and analyzing data for both State and local purposes than to concentrate on any limited approach.
 - 2. It is more efficient and cost-effective to build a common data analysis tool statewide than to work independently. Districts and the State can collaborate to improve the quality of their collection and analysis capabilities.

This project proceeded according to a development plan that allowed for a structured manner of defining project goals, soliciting competitive proposals, ensuring adequate training and project support, and adapting to project changes. This project plan included the following elements.

- A. *District Consultations.* The Department met with representative districts in person and via teleconference in early 2002 to outline the general needs, expectations, and criteria for a data analysis and reporting application.
- B. *Consultations with between DPI and ITD.* Representative from DPI and ITD met to discuss the merits of pursuing either a sole-source or RFP process for the data analysis project. It was the consensus of both parties to pursue a RFP process.
- C. *Business Case Development.* With the guidance of ITD, DPI developed a business case that summarized the need for and the expected cost benefits of pursuing a license contract with an outside contractor. A copy of the business case has been presented previously to the Information Technology Committee.
- D. *RFP Development.* DPI prepared a RFP, based on criteria identified in consultations with representative districts, outlining the technical requirements of the data analysis and reporting application. DPI consulted with TetraData Corporation, the firm selected by the Fargo District for a similar application, to aid in the technical aspects of the RFP. Following an internal review and after consulting ITD, the RFP was approved by DPI and released to prospective bidders. A copy of the RFP has been presented to the Chair of the Information Technology Committee.
- E. *Applications Review.* A committee of six DPI staff reviewed five applications in terms of the identified criteria. The first round of reviews eliminated three applications. A second round of interviews were conducted with representatives of the two remaining applicants. Following an independent scoring of all reviewers, the committee unanimously selected TetraData Corporation as presenting the best all-around proposal, considering both technical capacity and on-site support.
- F. *Contract Development.* With the assistance of ITD, DPI drafted a contract with TetraData to develop and support a statewide roll-out of the application. The contract was signed in October 2002.
- G. *EduTech Support.* The DPI entered into contract with EduTech to provide technical assistance to the project. EduTech provided direct contacts to schools and districts to discern project setup and training needs. EduTech provided first-line assistance with schools and districts and acted as a liaison with TetraData Corporation regarding technical issues, e.g., logon, warehouse structure, and training.
- H. *Statewide Student Identification.* As a corollary effect of establishing the statewide data warehouse, it became essential to develop a statewide student identification system in order to secure reliability within the warehouse structure. This need had been longstanding prior to this project; however, this project offered the best opportunity and the funding to develop this statewide student identification system.

- I. *Training.* TetraData Corporation provided for the project's training component. EduTech provided assistance throughout the project as an element of its liaison duties. Numerous training sessions occurred throughout the course of the project.
- J. *On-site Warehouse Consultations.* According to the development plan, thirteen data warehouses were established to accommodate the needs of the state's largest districts and their unique data needs, a single warehouse for the larger number of smaller districts and their more uniform needs, and the state in the aggregate. TetraData Corporation provided numerous on-site consultations with districts regarding the structure and implementation of their respective data warehouses.
- K. *Warehouse Training Documentation Applied to School Improvement.* The DPI contracted with Dr. Victoria Bernhardt, a nationally recognized specialist in data-driven decision-making and school improvement, to develop a guidebook on learning and using the statewide data warehouse for practical school improvement activities. This guidebook is available on the DPI website for ready access: <http://www.dpi.state.nd.us/resource/tetradata/index.shtm>.

This project has supported the stated objectives within the DPI's state technology plan, specifically the sections related to evaluation and continuous improvement.

Project Benefits

The project development plan identified several benefits to be derived from this statewide data warehouse. These benefits included the following.

- 1. *Data-driven Decision-making:* Educational literature identifies data-driven decision-making as among the most compelling challenges that face educators today. Schools, districts, and states are required to evidence true, supportable performance improvements. This expectation requires schools, districts, and states to provide for the capability to review and analyze performance data system-wide.
- 2. *Accountability:* State accreditation and federal reporting requirements based on individual student test data and other factors require schools, districts, and states to link student demographic and achievement data with school curriculum, teacher files, and various infrastructure databases.
- 3. *Support to Policymakers:* An integrated, statewide data warehouse provides policymakers with access to linked databases. Integrated data allow for the analysis of system-wide performance levels, capacity, programmatic initiatives.
- 4. *School and District Paperwork:* An integrated, statewide data warehouse simplifies the compilation and reporting of school, district, and state performance levels, thereby lessening school district paperwork and reporting processes.

5. *Cost Effective Statewide System:* An integrated, statewide data warehouse provides for a more robust linkage of databases and results in improved data analysis. All schools, districts, and the state have access to the total warehouse. Schools, districts, and the state contribute their respective databases and optimize data analysis tasks. The system provides for all warehouse needs and eliminates unnecessary duplication with individual school or district warehouses. Data warehouse costs are minimized and supported by the state.
6. *Valid and Reliable Reporting:* An integrated, statewide data warehouse enhances the prospects of conducting data analysis based on a system that compiles, processes, and reports data in a valid and reliable manner. Quality assurance measures are conducted according to set protocols. Schools, districts, and the state benefit from improved data collection, cleansing, and review.
7. *Optimizing Timeliness of Data Analysis:* An integrated, statewide data warehouse reduces time and effort required to compile various databases. Teachers and administrators can readily access databases that are linked by reliable identifiers. Less time is required to account for database variances, unclean data, inappropriate query designs, advanced analysis tools, and final report generation. The warehouse provides sufficient capacity and quality safeguards to improve the likelihood of desired results.
8. *Improved Communications:* An integrated, statewide data warehouse improves the likelihood of producing valid and reliable reports. The timeliness of queries, analysis, and reporting are enhanced. Reports are more easily generated as tables and graphs in order to make the information more accessible to a variety of users. Users are more apt to receive and understand reports that provide graphic presentations and supporting detail information.
9. *School Improvement Process:* Every school, district, and the state is required to undergo some form of school improvement activity based on performance data. An integrated, statewide data warehouse enhances the quality of generated school profiles. Data analysis is more apt to be appropriate when the process of querying and reporting are conducted according to reliable protocols. Program improvements are more apt to be appropriate when reliable reports are reviewed and understood.

Benefits Reviewed

1. *Data-driven Decision-making:* The statewide data warehouse links core data files (i.e., ACT scores, class, course, course grades, CTBS, district, school, dropouts, enrollment, graduates, state assessment, student, teacher) using confidential, unique student and staff identification numbers. The warehouse retrieves, compiles, organizes, analyzes, and reports the selected information according to user query directions. Reports are manipulated to allow flexibility for users. Reports are presented in tables and graphs to allow users to view data and conduct additional formula-based or logical arguments to arrive at meaningful results. These results allow users to base policy and programming

decisions on actual data. The warehouse evidences a fully functional, robust architecture capable of simple or highly complex queries and report generation.

2. *Accountability:* The statewide data warehouse is designed to allow school personnel to access and analyze the data for which each school and school district is being held accountable. The warehouse contains all the primary data used in federal accountability reporting and a substantial degree of data used in state accountability reporting. The warehouse is fully capable of incorporating additional data files to establish broader reporting. The DPI has identified future expansion to the data files, including adding a family of finance files, teacher assignment files, and student inclusion files. The warehouse is capable of unlimited expansion.
3. *Support to Policymakers:* The DPI has received comments from district personnel that the warehouse has been used for the analysis of some programmatic and policy issues. The DPI used the warehouse for certain data discovery activity during the recent legislative session. The use of the warehouse for such activities has not been surveyed; it is expected from anecdotal evidence that the warehouse is still underused for policy review at this early point in its implementation.
4. *School and District Paperwork:* The statewide data warehouse allows schools and districts to intersect different databases and to generate reports. The warehouse minimizes effort and time to conduct data intensive background research. The warehouse minimizes the work required to generate meaningful reports, both in tables and graphs. The warehouse evidences a robust architecture that processes extensive, complex queries quickly and accurately.
5. *Cost Effective Statewide System:* The statewide data warehouse provides for the data analysis needs of all schools, districts, and the state. The warehouse is provided for within one contract and supported through DPI federal discretionary funding. There has been no cost to any of the school districts for the implementation or use of the statewide data warehouse.
6. *Valid and Reliable Reporting:* The schools, districts, and the state populate the statewide data warehouse with their respective data based on established quality assurance protocols, when properly conducted. Each school, district, and the state assume responsibility for the accuracy and currency of their respective data. Protocols exist that either require or encourage data cleansing to ensure validity and reliability. All data queries and reports, when designed comparably, will produce comparable results. Many pre-programmed queries are provided to ensure consistent, accurate queries and reports. The DPI have provided common training materials, designed around the school improvement process, that enhance the prospects of reliable query and report design. Schools, districts, and the state have the capability of producing common performance report cards that are valid and reliable consistently.
7. *Optimizing Timeliness of Data Analysis:* The statewide data warehouse provides for efficient and accurate queries with proper training. The warehouse is a sophisticated tool that does require a minimal level of training for basic queries and more advanced training for complex queries and reports. With a reasonable

level of training, any user may produce respectable reports that demonstrate higher levels of data analysis. Training, however, is required. A sophisticated, robust tool such as the statewide data warehouse does require familiarity and use to acquire user proficiency. The current warehouse only provides for biannual data uploads, which limits the timely analysis of fluid data. The DPI has prepared a contract with TetraData Corporation to provide for data-on-demand functionality, which would allow for up to daily uploads and enhanced capability.

8. *Improved Communications:* For the purposes of general accountability and the determination, the DPI references data from the Department's Online Reporting System and analyzes the data according to dedicated rules. Districts that have used the statewide data warehouse have reported using it for general programmatic purposes. These reports indicate that turn-around time has been improved and a general reduction in effort is required. There has been no survey conducted to assess the various programmatic or policy areas where the warehouse has been used.
9. *School Improvement Process:* The State and school districts can now conduct longitudinal analysis, cohort tracking, and disaggregated and analysis of individual student performance. Emerging anecdotal accounts indicate that some schools, districts, and school consortia have begun to use the warehouse as a means to prepare reports, including profiles, for school improvement efforts. There has been no survey conducted to assess the various uses of the warehouse for school improvement.

Lessons Learned

1. Pre-populating the statewide data warehouse with data collected on the Department's Online Reporting System (ORS) saves time and improves the overall accuracy of the data. Data collected on the ORS is used for general accountability reporting and payments. Populating data from the ORS into the statewide warehouse builds efficiency by accessing a primary data source. The ORS contributes a substantial percentage of data collected for all schools and constitutes a common core collection of data used statewide. This common core data ensures greater reliability and comparability for generated reports. Using the ORS as a primary data sources saves time and limits unnecessary duplication in the collection process for the school districts.
2. The statewide data warehouse has heightened the attention of school personnel on the relative quality of their individual databases. Throughout the project, especially in the earliest phases of data collection, school personnel reviewed the data compiled in the warehouse against the schools' primary data sources. School personnel have generally encountered various levels of "unclean" data that have originated in their own data collection processes. For school personnel who have attended to this matter, there have been reports of greater vigilance in the data collection process and commensurate improvements in the quality of overall data. The cleaning of data has arisen as a primary need statewide.
3. The initial training on the statewide data warehouse occurred too early in the project. The project's original design was to introduce prospective users early in

the development phase to the design and use of the warehouse. This original scheduling was intended to improve the investment of users into the data collection and cleansing process and to make the project's goals more meaningful. During this training, prospective users referenced artificial data while their own data underwent cleansing. The actual rollout of the project resulted in the data cleansing process taking more time than originally anticipated which in turn resulted in delays in the actual use of the warehouse. This resulted in the diminution of both user skills and the training's effect. Experience indicates that it is better for users to train using their own warehouse and data instead of a 'dummy' warehouse.

4. In the earliest phases of the project, much of the project's activity surrounded the compilation and cleansing of data. This activity did not optimize the benefits of contractors meeting regularly with various users, especially users in smaller districts. Communication and 'buy in' of the project improved once the project's contractor began monthly field visitations to offer general technical assistance and warehouse applications. The state's largest districts received a substantial level of attention throughout the setup of their respective warehouses, given the size of their respective individual data systems. This attention resulted in the largest districts generally having sufficient time to consult with the project's contractor regarding technical needs and data collection protocols. Larger districts, by the nature of their size and needs, received more attention than small districts. Smaller districts received the benefit of a general statewide warehouse; however, given their generally smaller data systems, they did not require or receive much attention. All smaller districts were contacted to assess and attend to their data needs; however, most elected not to pursue this invitation.
5. The DPI contracted with EduTech to provide overall Level 1 technical support for school districts, especially smaller school districts, regarding all aspects of their data compilation, cleansing, and reporting needs. EduTech managed the overall support for the smaller districts' multi-district warehouse. EduTech reported directly to the project's contractor regarding any technical issues and to the DPI on overall project flow. EduTech staff also shadowed the project contractor's staff on various field trips and training opportunities. The appointed EduTech staff did not have previous experience in the use of a warehouse for education improvement; this did result in a reduction in effectiveness.
6. The project contractor's staff built and loaded the warehouses with standard queries and reports that were unique to the design of the statewide data warehouse. These queries and reports were based on requests from various district warehouse users. These queries and reports were designed to be universally applicable to any school, district, or state need. Some districts and consortia specifically requested and received assistance in the generation of reports that were customized for their own unique needs. The DPI will work to update and increase the number of these template queries and reports to address a wider scope of education improvement issues.
7. The development of individual student identification codes early in the design of the system has enhanced the overall reliability and functionality of the Department's Online Reporting System and the statewide data warehouse. The development of the individual student identification codes marks a substantial

improvement to data use statewide and a benchmark accomplishment of this project.

8. This project has illuminated the data management practices in use statewide. The state's schools and districts use a variety of electronic data management tools and some have no electronic data management at all. During the various phases of this project, it became apparent that many schools and districts had little or no data quality assurances built into their data management processes. In many instances, data cleansing has been managed more as a luxury than as a programmatic necessity. Some school personnel could not recognize or denied their own data when it was presented to them in a disaggregated form. The current state of "data culture" varies greatly across school districts. By their size, larger school districts are more apt to perform some level of data review and to improve their data collection processes than smaller districts. Increased performance accountability with its ready reference to primary data sources has further heightened the need to attend to data collection and analysis practices.
9. Some school and district personnel raised doubts regarding the need for or use of a statewide data warehouse. Other comments from smaller schools or districts indicate that some staffs had a sufficient grasp of their own students and their data. Still other comments indicate that some perceived little benefit in comparing their respective data against data compiled statewide. Some staffs have indicated their preference to conduct school improvement in terms of anecdotal evidence instead of objective data analysis. There appears to be a wide variance statewide in the use or value of data collection and analysis practices.
10. The need exists to upload data on a regular, perhaps daily, basis. Users of the statewide data warehouse requested that the warehouse allow for more than the project's original biannual updates. This recommendation indicated that users were recognizing the value in using the most current data possible. The DPI proceeded to negotiate a data-on-demand service with the project's contractor for the next contract period. Data-on-demand would allow for up to daily uploads of data from the Department's Online Reporting System, local school districts' individual data systems, and other outside data sources. This function would mark a substantial and needed improvement in the compilation and use of real-time data.
11. The statewide data warehouse is a robust, sophisticated data management tool. It allows for a wide variety of data analysis functions and offers the capacity for unlimited expansion. Although its user interface is manageable with limited training, some school personnel have requested a friendlier interface to ease its use. The DPI has negotiated with the project contractor for the introduction of a newer user interface in 2006. This newer interface is designed to allow more users to access the power of the warehouse without extensive training.

Customer Response

At the request of the DPI, EduTech conducted a user survey of the statewide data warehouse in January 2005.

Comments from respondents identified two issues that seemed of primary concern to the smaller schools within the Multi-District warehouse: (1) the value of the warehouse and (2) the training to support the warehouse. These two issues are addressed in subsections 3 and 9 in the preceding section. The Multi-District warehouse had been managed by EduTech.

The survey responses from the larger districts were generally favorable. Respondents supported the goals and uses of the warehouse, although respondents acknowledge that pressing daily responsibilities and time conflicts limited the use of the warehouse. These statements are similar to those from districts related to other general school improvement activities and obligations. In addition, several school districts were implementing Power School, which required a great deal of time and attention from their technology personnel. Since the time of the survey, Dickinson and Mandan have become avid users of the system, and Williston and Fargo have requested uploads of their data. The DPI has received additional requests for the uploading of state assessment and other ORS data during the current interim service contract with the project contractor.

The DPI has been engaged in discussions with representatives of four larger school districts regarding the future goals of a warehouse and the upgrading of functionality for data-on-demand and the user interface.